

基于 SOA 的数字图书馆 OLE

谢建琳

【摘要】 基于 SOA 开放框架的数字图书馆 OLE, 旨在设计新一代的更开放式图书馆自动化系统, 更方便快捷地满足 21 世纪多变、复杂的用户需求。SOA 为构建大规模、资源可重用、松散耦合分布式系统提供了一个开放的体系结构。为了实现 SOA 架构, 其部署实施的基础架构必须支持相应的系列原则。SOA 是让 IT 更加关注于业务流程而非底层 IT 基础结构, 从而获得竞争优势的更高级别的应用程序开发架构, 因此, 对需要使用信息技术解决的业务问题上很有益处。

【关键词】 OLE SOA 面向服务 数字图书馆 体系架构

Abstract: The Digital Library OLE Based on SOA aims to design a more open library automation system which is comparatively convenient, so as to meet the changeable and complex needs brought up by the clients of 21st century. SOA provides an open structure for building a large recycling system that loosely coupled distributed. In order to achieve the aim, the constructing infrastructure must stick to the corresponding principles. And the essential of SOA is to make IT focus on business process instead of infrastructure, then to gain the higher-level application of development architecture with its competitive superiority. Therefore, SOA is useful to those problems which require info-tech to solve

Key words: OLE SOA services-oriented digital library architecture

1 引言

随着数字图书馆 DLS 的日益增多, 人们发现必须访问许多 DLS 才能找到所需要的资料, 给学习、工作带来诸多不便。人们希望图书馆能向用户提供一个分布信息源无缝集成的视图, 并提供个性化的服务。OLE 计划正是基于此目的提出的。OLE (Open Library Environment) 计划是指由美仑基金会提供 475, 700 美元给杜克大学图书馆, 旨在设计新一代的更开放式图书馆自动化系统, 方便快捷地满足 21 世纪更多变、更复杂的用户需求^[1]。基于这一点, 本文提出了基于 SOA 的开放框架实现图书馆的 OLE。

2 什么是 SOA

面向服务的体系架构 (service-oriented architecture, SOA) 是一种粗粒度、松耦合的服务结构。SOA 是服务的集合, 服务是最核心的抽象手段, 业务被划分 (组件化) 为一系列粗粒度的业务服务和业务流程^[2]。其目的是最大限度地重用应用程序中立型的服务以提高 IT 适应性和效率。它可以根据需求通过网络对松耦合的粗粒度应用组件进行分布式部署、组合和使用。服务层是 SOA 的基础, 可以直接被应用

调用, 从而有效控制系统中与软件代理交互的人为依赖性^[3]。SOA 的关键是“服务”的概念, W3C 将服务定义为: “服务提供完成一组工作, 为服务使用者交付所需要的最终结果。最终结果通常会使使用者的状态发生变化, 但也可能使提供者的状态改变, 或者双方都产生变化”。在 SOA 中, 服务是最基本的单元, 提供了服务提供者的语义特征、接口特征和服务的各种非功能特征, 通过标准协议和数据格式进行通信。目前广泛采用 XML 作为标准的数据交换格式。

3 SOA 的体系结构

图 1 是 SOA 的体系结构图。其中, 服务提供者 and 请求者之间能够进行动态绑定和直接交互, 实现企业应用的逻辑。SOA 强调服务的封装、重用以及服务之间的互操作性, 通过“描述-发布-发现-交互”的一系列标准操作, 为构建大规模、资源可重用、松散耦合分布式系统提供了一个开放的体系结构。

图 2 给出了基于 SOA 的 DLS 开放框架简化模型, 该框架分为四个区: 用户区、应用区、注册区、服务使用者。

服务区是服务提供者的集合, 利用底层的组件实

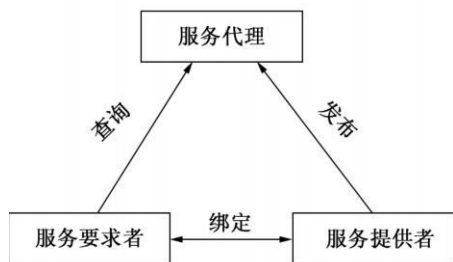


图1 SOA体系结构

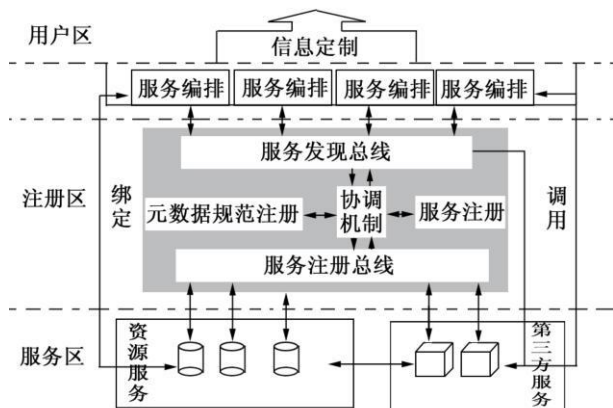


图2 SOA框架结构

现服务，并屏蔽底层平台的差异，提供一致的服务组件模型。服务区主要包括 DLS 的资源服务、以及第三方服务（开放搜索、认证、付费、元数据转换等）。第三方服务有时需要使用 DLS 的资源服务，因此第三方服务既是服务的提供者也是使用者。这里主要讨论资源服务，因为它是该框架的基础，将 DLS 的内部资源以服务的形式体现。它主要包括以下模块：元数据的收割模块、服务注册模块、服务实现模块。其中元数据的收割模块参考 OAI-PMH 框架^[4]，实现对图书馆数字资源的元数据集成和整合。服务注册模块，将实现的服务接口和收割后的元数据规范发布到服务代理者，提供服务使用者查询。服务实现模块将框架中的服务提供者逻辑组件封装成服务或者直接用服务组件模型实现资源服务。

注册区是注册操作入口站点的集合，实现代理者功能，提供一种特殊服务，这种服务实现通用的注册、发现机制^[5]。该机制为框架中资源服务、第三方服务和资源元模型的开放集成和互操作提供了有效的保证。任何注册的服务都可以被第三方发现，并集成到第三方系统使用（当然需要通过服务提供者的认证）。在服务代理者中除了服务的注册外，为了实现元数据互操作还应包括资源元数据模型的注册，涉及标识符规范、元数据描述规范、资源集合描述规范、知识体系描述规范、管理规则描述规范等注册。

应用区是服务使用者的集合，是资源服务和第三

方服务的使用者，但同时也向用户区提供应用服务，该区主要任务是实现服务的编排，按照用户的需要、基于用户的知识体系或者业务流程来定制个性化信息服务系统^[6]。应用区中的服务使用者查询服务代理者上注册的服务描述信息，然后绑定服务接口。由于服务完全屏蔽了底层的实现细节，无论采用 J2EE、NET 或者传统面向 API 的系统，服务组件模型都可以实现良好的封装，因而面向服务的集成完全是开放的解决了图书馆的异构集成、互操作问题。

用户区实际上是一个接入终端，可以是浏览器，或者用户自己定制的应用程序，发出应用请求，比如智能检索、订阅电子资源等。

框架的运行机制主要体现在 SOA 的协作上：服务提供者通过服务注册总线将服务信息和元数据规范信息注册到服务代理者的信息库中；服务使用者通过服务发现总线向服务代理者查询服务信息，然后根据返回的信息向服务提供者获取服务的描述，进而绑定服务；根据用户的请求，服务使用者（包括第三方服务和应用区中的业务流程）调用服务。

4 SOA 的架构原则

为了实现 SOA 架构，部署实施的基础架构必须支持 SOA 原则。这些原则是：

明确边界：服务被调用时，与实现其功能相关的内容都应被传递过来。对服务的所有访问都应该通过公共接口进行。调用服务时，非隐含的假设是必须的。作为通用的模式，服务调用不应依赖于共享的上下文，而应被作为无状态的模块。服务调用是一个具有业务逻辑效果的行为，可能有大量的资源开销，并且导致一系列不同于本地方法调用和远程过程调用的错误。服务的调用绝非远程过程调用。服务的使用和提供应该尽可能地简单，因此与服务间的交互没必要被隐藏得太多。在 SOA 中，服务发送和接收的消息、服务契约以及服务本身都应当是最好的构件。一个明确的接口会封装服务的内在实现，而服务通过该接口发布自己的功能，与服务交互是一个具体的行为，它依赖于服务使用者和提供者之间消息的传递。

共享契约和架构：基于一份服务描述（一份契约），服务使用者和服务提供者都可以获得使用或提供服务的全部所需。根据松耦合原则，服务提供者不能依靠服务使用者来重用那些依赖于使用者环境的代码。毕竟，服务使用者可能使用不同的开发环境和运行环境。这条原则给 SOA 体系中所能交换的数据加上了严格的限制。理想的情况是，数据以符合一种或多种模式的 XML 文档形式被交换，因为这种方式可应用于任何你能想到的编程环境。

策略驱动: 为了与服务交互, 必须满足不同的要求: 提供者提供的功能、语法和语义必须适应使用者的需求; 技术能力与需要必须匹配。为了使尽可能多的形形色色的使用者能对服务进行访问, 一种策略机制已经被作为 SOA 工具集的一部分引入了。在服务接口对功能进行描述的同时, 策略对不同的, 非功能性的能力和需求进行了指定。

自治: 与明确边界原则相关, 服务自治意味着, 接口成为服务与外界联系的唯一方式, 至少从 SOA 的角度来看是这样的。要注意服务的运行环境一定是可变的。服务能够被彼此独立的修改、部署、发布新版本和管理。服务提供者不能寄希望于服务使用者, 期望它们依靠自己的能力迅速适应新版本的服务。

采用可传输的协议格式, 而非 API: 服务通常采用协议格式来发布, 协议格式应该是明确的、可传输的并且被服务所支持的。为保证一个服务最大程度的可访问性 (及长期的可用性), 只要交互过程遵守为该服务定义的策略, 那么由任何依照服务接口进行信息交换的平台都可以访问该服务。

面向文档: 服务交互时, 数据是以文档的形式来传递的。文档是一个被明确模块化的, 有层次结构的数据容器。面向文档的一个重要特征就是自描述。最理想的情况下, 文档是对现实世界中的文件 (如订单、发票或帐单) 的建模。文档应该被设计来确保它在问题域的上下文中发挥作用, 这意味着它们可能应用于一个或更多的服务。

松耦合: 它是 SOA 中一个很重要的概念, 一个系统可以在多个维度表现为松耦合或紧耦合, 它依赖于具体的要求和上下文, 系统可能会在一些维度是松耦合, 在另一些维度是紧耦合。这些维度包括: 时间、位置、类型、版本、基数、接口。创造一个满足以上所有维度的松耦合系统, 既不可行, 也没必要。不同类型的服务要做不同的取舍^[7]。

遵循标准: 一个 SOA 应用中应遵循的一个关键原则是, 信赖标准而非专有的 API 和格式。标准存在于技术方面, 如数据格式、元数据、传输协议; 也存在于业务层面, 如文档的类型。应注意何种情况用何种标准, 标准的一个重要方面是它的可接受性。

元数据驱动: SOA 中所有的元数据对象都需要被按照一种方式储存起来, 这种方式将确保元数据对象能够在设计和运行时被发现、检索和解释。元数据对象包括对服务接口、参与者、端点和绑定信息、组织单元和职责、文档类型或模式、使用者或提供者关系等的描述。这些对象的用途应当是被代码自动生成或者解释, 成为服务和参与者生命周期的一部分。

5 面向服务架构的优点

面向服务架构是让 IT 更加关注于业务流程而非底层 IT 基础结构, 从而获得竞争优势的更高级别的应用程序开发架构。SOA 对需要使用信息技术解决的业务问题 (包括希望减少冗余架构、创建跨用户和员工系统的公共业务接口、需要基于角色和工作流对用户提供个性化信息业务、希望通过 Internet 实现经由移动设备的访问来提升客户服务的组织) 很有益处。具体表现在:

效率: 将业务流程从“烟囱”状的、重复的流程向维护成本较低的高度利用、共享服务应用转变;

响应: 迅速适应和传送关键业务服务来满足市场需求, 为用户提供更高水准的服务;

适应性: 更高效地转入转出让整个业务变得复杂性和难度更小, 达到节约时间和资金的目的;

复杂性降低: 基于标准的兼容性, 与点到点的集成相比降低了复杂性;

重用增加: 通过重用以前开发和部署的共享服务, 实现了更有效的应用程序、项目开发和交付;

遗留集成: 用作可重用服务的遗留应用程序降低了维护和集成的成本。

6 结语

实现开放式数字图书馆是当今数字图书馆界和软件工程界一直关注的课题, 也是 OLE 计划的目的。本文提出的基于 SOA 架构的目的在于: 在众多分布、异构和自主的信息资源系统组成的开放环境中, 数字图书馆可以敏捷地根据具体的信息用户的个性化需求, 发现、解析和调用所需要的资源和服务, 按照个性化的服务流程和业务逻辑将信息资源或服务灵活组织起来构成新的服务。但对于实际应用的系统必须解决以下几个问题:

性能: 开放框架使用的是动态绑定, 服务的执行是非常低效的, 通过缓存 bindingkey 来提高服务执行性能是一条可行的方法。

数据同步: 对于开放框架而言对跨区域的 DLS 代理中必然包括多个操作入口点, 如何保持操作入口点之间的信息同步也是一个问题。

敏捷性: 开放框架能够对需求变化迅速做出反应, 能够快速调整信息系统的结构布局。

信息交互性: 各节点物理上是分布的, 平台上是异构的, 各节点间信息能正确流动, 各节点间信息语义要一致, 需要标准、通用的信息交换接口, 便于信息的无缝连接与传递。

安全: 如何构建一个有效安全机制保证服务提供

(下转第 22 页)

应该从“控制导向型”转向“保持距离关系型”，并形成以委托代理关系为基本框架、以契约制为治理机制、以提高知识资源战略使命性和效率性为主要目标的新治理模式。

其次，知识资源战略实施的调查与规划先行策略，即在知识资源战略实施之前，必须有针对性地调查分析和规划设计。美国在 2000 年组织大规模的数字信息资源调查，发布包括 Louis Pitschmann 的《公益网络信息资源可持续建设规划》报告，Timoth Jewell《商用数字信息资源选取与保存理论与实践》报告，以及 Abby Smith《数字信息资源战略规划报告》在内的有广泛影响的研究文献，为数字信息资源建设提供了丰富的实践和理论支持。国内知识资源战略实施的调查与规划工作应主要包括三个层面，即国家层面、行业层面和社会公众层面。

第三，知识资源战略实施的法律保障策略，即知识资源战略实施过程、资金投入、责任权利划分等问题应以法律法规的形式予以保障。美国政府在 1999 年修改后的《联邦信息资源管理政策》中详细规定了信息共享和数字信息资源开发实施细则。由于国内具体法律制定程序的限制，及时颁布相关法规和政策性文件的可能性不大，但从长远考虑，国家层面的知识资源战略实施必须有政策、法律框架的支撑。从研究的角度讲，现阶段知识资源战略实施的法律框架和政策框架研究应该与调查规划工作同步展开。

第四，知识资源战略实施的战略选择策略，即确定先开展哪些战略。由于国家战略层面的知识资源区别于信息资源的意义主要体现在民族性、应急性、记忆性和知识创造基础的价值，创新型国家知识资源战略实施首选的项目应该是能够体现知识资源对于创新

国家建设具有记忆价值、民族价值和应急价值的，例如汶川地震后的地理知识资源建设，当地的非物质文化遗产建设，以及具有民族性的知识、技能和文化资源建设等。另一方面，具有科学研究基础价值的知识资源和面向社会公众的普适性知识资源也应优先建设。

第五，知识资源战略实施的开放性和共享性策略，即知识资源战略实施的最终目的是能够为各种用户提供获取渠道，并能保证不同知识资源项目之间可以无缝衔接。有关方面估计我国科学数据总量占世界的 10%，每年发表的科技文献量约占全球的 1/8，但数据共享所产生的效益则只占全世界的 0.1%。大量资源处于闲置状态，科学数据获取方面的障碍造成国家在文献资源领域投资的效益得不到发挥，已经阻碍了我国科技创新活动。因此，知识资源战略实施的最终目的是保障用户有资源可获取，并有渠道能获取，以此发挥知识资源在创新型国家建设中的支撑作用。

注释

- [1] 宗诚，马海群．数字信息资源开发利用的宏观战略与规划研究．情报资料工作，2008 (4): 65-68
- [2] 孙建军，柯青，成颖．基于系统观的国家数字信息资源战略环境分析方法及规划模式．图书情报工作，2007 (11): 79-82
- [3] 蒋永福．论图书馆治理．图书馆论坛，2008 (12): 50-55
- [4] 格里斯·托克，华夏风译．作为理论的治理：五个论点．国际社会科学：中文版，1999 (1): 19-29
- [5] 王黎．治理视野下的多元主体之间的整合—市场、政府和 NPO 的合作伙伴关系．湘潮，2007 (6): 64-66
- [6] 洪秋兰．公共事业组织知识管理的实施模式构建．图书情报工作，2007 (10): 48-51

王 平 郑州大学信息管理系。

(上接第 54 页)

者和使用者安全、高效地使用信息服务。

SOA 作为新型的数字图书馆体系架构方案，只有解决好以上问题，完善整个框架，才能实现将信息技术从传统的“以系统为中心”转向“以服务为中心”；遵循“业务驱动服务，服务驱动技术”的新理念，才能实现“以用户为中心”的业务流程敏捷应变。

注释

- [1] OLE project <http://www.librarytechnology.org>, 2008-09
- [2] 毛新生．SOA 原理·方法·实践．北京：电子工业出版社，2007: 3-4

- [3] 面向服务的体系结构概述．<http://www900.ibm.com/develop/works/cn/Webservices/ws-ovsoa/index.html>, 2004-06
- [4] Liu Xiao Ming Repository synchronization in the OAI framework// Digital Libraries Proceedings 2003 Joint Conference 2003: 27-31
- [5] UDDI v3.0 <http://www.oasis-open.org/news>
- [6] PELTZ C Web services orchestration and choreography. Computer, 2003, 36 (10): 46-52
- [7] 邹凯，陈添源．一种面向服务的数字图书馆体系架构．情报资料工作，2007 (2): 60-63

谢建琳 福州大学图书馆。